Agricultural and Environmental Surveys

Roberto Benedetti
University of Chieti-Pescara, Italy

Federica Piersimoni
ISTAT, Italian National Statistical Institute, Rome, Italy

Francesco Pantalone
University of Perugia, Italy

Contents

- The use of GIS
  - General features
  - Application to agricultural and environmental statistics
- The use of Remote Sensing
  - General features
  - Application to agricultural and environmental statistics
- Setting up the frames
- Stratification
- An overview of Some existing Agricultural and Environmental surveys based on Area, Point or Multiple Frames
  - ESYRCE (Encuesta sobre Superficies y Rendimientos de Cultivos)
  - LUCAS (Land Use and Cover Area frame Survey)
  - TERUTI (Utilisation du territoire)
  - JAS (June Agricultural Survey)
  - AGRIT

GIS

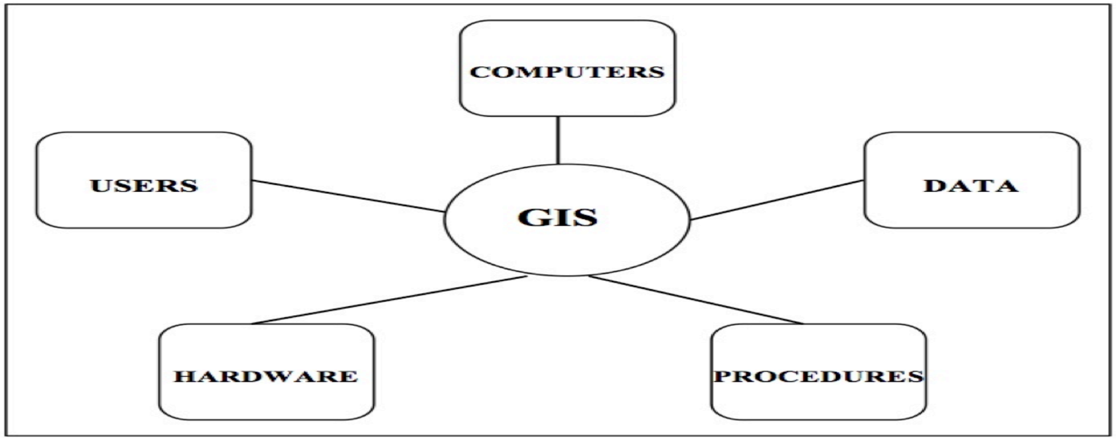
Geographical Information System (GIS).

Spatial database aimed at the operational and efficient management of data related to a space.

Several definitions in literature. In essence, we can refer to a GIS as a computerized information system that integrates data, hardware, software and the Global Positioning System (GPS) to assist a researcher in the analysis and display of geographically referenced information.

GIS is capable of handling the positions of elements in a territory and provides query and visualization software components.

GIS



Generally, a GIS incorporates five key elements: hardware, software, users, procedures, and data.

- *Hardware*: the computer that operates.
- *Software*: Key components of the software are tools for the inclusion and the manipulation of geographic information, the database management system, tools for research, analysis, and data visualization.
- *Users:* they manage the system and develop projects.
- *Procedures*: the system has to be structured in a operational coherent way.
- *Data*: the most important component constituted by four elements: the geographical position, the attributes, the spatial relationships, and the time.

GIS

The analysis represents the core of any GIS system because allows us to visualize the patterns and relationships of geographical data.

Current methods can be very simple such as a map of the theme under investigation, or more complex models such as those that simulate the real world by combining many layers of data.

There exist an extensive number of methods for analyzing geo-referenced data. In particular, these techniques are grouped into two classes: *spatial analysis* and *spatial statistics*.

GIS

Spatial analysis studies real-world processes (typically uses a GIS), providing information about the real world and the current situation in terms of specific areas and features, and changes in situations or trends. It is generally characterized by three elements:

- *attribute data*, *a-spatial* components that provide the descriptive information of the spatial data;
- *geographical location*, (*spatial component*) is defined using coordinates expressed in a given system of reference (such as latitude and longitude);
- *topology,* defined as the spatial relationship between map features (representation of a real-world object on a map) and concerning some features required for spatial analysis as adjacency, containment, connectivity, and intersection

A GIS provides the tools to deal with all three elements of spatial data.

GIS

Spatial statistics concerns methods for understanding data distributed in a space, where position and distance have meaning.

When considering the spatial statistics, the analysis capabilities of a GIS system can be substantially improved by combining it with *R*.

The use of GIS in spatial statistics could lead to a number of advantages.

- Recent crop surveys are based on GIS and area sampling methods. In fact, the sampling units in crop surveys are based on an area frame obtained using geographical areas such as villages, cities, and regions. Census data, survey data, and satellite images are all integrated into a GIS.
- Statistical precision can be improved by including agricultural or other environmental characteristics into the prediction models.

GIS

- The visual nature of maps may highlight hidden relationships that are very important in a standard regression analysis.
- A GIS can also be used to create more useful data for rural development policy makers.
- Small area estimations combined with geo-coded data and GIS can provide a different view of the poverty distribution at a sub-region level.
- Possibility to improve the accuracy of crop area estimates by incorporating the effect of spatial dependencies through an integrated application of remote sensing technologies and GISs.

References

- Aranoff S (1991). Geographic information systems: A management perspective. 2nd edition. WDL Publications.
- Burrough (1986). Principles of geographical information systems for land resources assessment, Clarendon Press, Oxford.
- Chou YH (1997). Exploring spatial analysis in geographic information systems. OnWorld Press.
- Cowen DJ (1990). GIS versus CAD versus DBMS: what are the differences?. In: Introductory readings in Geographic information systems, Peuquet DJ, Marble DF (eds), Taylor & Francis, London, pp. 52-61,
- ESRI (1990). Understanding GIS: The ARC/INFO method, Redlands, CA, USA.
- Köhl M, Magnussen S, Marchetti M (2006). Sampling methods, remote sensing and GIS multiresource forest inventory, Springer-Verlag Berlin Heidelberg.
- Fischer MM, Nijkamp P (2013). Handbook of regional science. Springer-Verlag, Berlin.
- Goodchild MF (1992). Geographical data modeling. Computers and Geosciences, 18: 401-408.

References

- Goodchild MF (2007). The Morris Hansen Lecture 2006. Statistical perspectives on spatial social science. Journal of Official Statistics, 23: 269–283.
- Goodchild MF, Longley PA (2013). The practice of geographic information science. In: Handbook of regional science, Fischer MM, Nijkamp P (eds), pp. 1107-1122.
- LeSage J, Pace K (2009). Introduction to spatial econometrics, Chapman & Hall/CRC, Boca Raton, FL.
- Liu J, Mason P (2009). Essential image processing and GIS for remote sensing. Wiley-Blackwell.
- Neteler M, Mitasova H (2008). Open source GIS: A GRASS GIS approach. Third Edition. Springer Scince + Business Media, LLC, New York.
- O'Sullivan D, Unwin DJ (2010). Geographic information analysis. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Shekhar S, Chawla S, Ravada S, Fetterer A, Liu X, Lu C (1999). Spatial databases - Accomplishments and research needs, IEEE - Transactions on Knowledge and Data Engineering, 11: 45-55.

Remote Sensing

Remote Sensing is defined as the technique for deriving information about the Earth's surface and estimating geo-bio-physical properties using electromagnetic radiation.

The data are acquired without physical contact with the Earth. This process involves making observations using sensors (i.e., cameras, scanners, radiometer, radar, and so on) mounted on platforms (i.e., aircraft and satellites) that are a considerable height from the Earth's surface, and recording the observations on a suitable medium (i.e., images on photographic films and videotapes, or digital data on magnetic tapes). Then, the observation data are typically stored and manipulated using computers.

Remote Sensing

RS system usually consists of:

- a platform, typically a satellite or aircraft;
- a navigation device, which establishes the location of the platform and the land area under investigation;
- one or more sensors;
- a module for data processing and interpretation (the interpreter can be a human or an automated system that supervises the whole operation and platform).

RS systems can be active (radar and laser emit their own electromagnetic radiation) or passive (electromagnetic waves that come from external sources of energy like the sun).

Remote Sensing

- When plants receive electromagnetic energy from the sun, there are three different possibilities. Depending on the wavelength of the energy and the plant's features, the energy will be reflected, absorbed, or transmitted. The differences in leaf colors, textures, and shapes determine how much energy will be reflected, absorbed, or transmitted and this will determine the spectral signatures of individual plants.

- Reflectance measured by a sensor can be considered as a proxy variable of some biophysical phenomena such as the geographical coordinates (x, y) of an object, the temperature, the color, or the moisture content of the soil and vegetation. These covariates are often called direct variables.

- We can derive some hybrid variables by simultaneously analyzing several biophysical variables. For instance, it is possible to determine the stress of a plant by considering its absorption characteristics, temperature, and moisture content. This represents a hybrid variable.

Remote Sensing

- Now RS uses special sensors that allow to investigate phenomena, even using infrared and microwave bands.

- Multispectral detection may provide new information that is not obtainable with visible spectrum methods, for example, infrared sensors measure the thermal emission of an object, and these temperatures can represent important parameters. From a statistical point of view, this information represents a typical example of a multivariate data set.

Remote Sensing

- A digital image is an extremely useful way to synthesize and display a large amount of data that could not otherwise be analyzed.
- The output of a digital image is typically a square grid divided into smaller boxes, also known as square *pixels* (i.e., picture elements). The element associated with each pixel corresponds to a numeric value that represents a gray level and describes the spectral luminosity of a specific area of the scene

Remote Sensing

RS techniques are widely used in agriculture and agronomy. In fact, remotely sensed images provide a spatial coverage of a field, and can be used as a proxy for measuring crop and soil attribute. It is really needed in agriculture monitoring due to some challenges (that are not found in other economic sectors)

- agricultural production heavily depends on seasonal patterns related to the life cycle of crops
- production varies according to the physical landscape (i.e., soil type), climatic conditions, and agricultural management practices
- agricultural variables vary substantially over space and time.

For these reasons, agricultural monitoring systems must be timely. RS has many advantages in that it can significantly help to address these needs, it is appropriate for collecting information over large areas, and can have a high revisit frequency.

Remote Sensing

- Spectral response and crop type do not have a one-to-one correspondence. In fact, the radiometric response of the same crop in different conditions can vary across the pixels of an image. A more appropriate approach is to consider the spectral response of a crop as a function of the probability distribution of its spectral reflectance.

- Satellite and/or aerial RS technology combined with *in-situ* observations has become an important technique for improving the present systems of acquiring and generating agricultural and resource data.

- RS has been increasingly considered for developing standardized, faster, and possibly cheaper methods for agricultural statistics.

Remote Sensing

- RS techniques can represent an appropriate support for particular problems in agricultural surveys such as data reliability, incomplete sample frame and sample size, unit selection, area measurement, non-sampling errors, gaps in geographical coverage, and non-availability of statistics at a disaggregated level.

- RS can be appropriately used at the design level.

- Remotely sensed images provide an overview of the area under investigation, and are useful when constructing the spatial reference frame.

- Classified satellite images can be used as auxiliary variables to improve the precision of ground survey estimates, generally with a regression or a calibration estimator.

- The remotely sensed information can also represent an auxiliary variable for small area estimation procedures.

References

- Arbia G, Benedetti R, Espa G (1999). Contextual classification in image analysis: An assessment of accuracy of ICM. Computational Statistics & Data Analysis, 30: 443-455.
- Atzberger C (2013). Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. Remote Sensing, 5: 949-981.
- Ball GH, Hall DJ (1965). Isodata: A method of data analysis and pattern classification. Stanford Research Institute, Menlo Park, United States.
- Becker-Reshef I, Justice C, Sullivan M, Vermote E, Tucker C, Anyamba A, Small J, Pak E, Masuoka E, Schmaltz J, Hansen M, Pittman K, Birkett C, Williams D, Reynolds C, Doorn B (2010). Monitoring global croplands with coarse resolution earth observations: The global agriculture monitoring (GLAM) project. Remote Sensing, 2: 1589-1609
- Benedetti R, Rossini P (1993). On the use of NDVI profiles as a tool for agricultural statistics: The case study of wheat yield estimate and forecast in Emilia Romagna. Remote Sensing of Environment, 326: 311–326.
- Benedetti R, Rossini P, Taddei R (1994). Vegetation classification in the middle Mediterranean area by satellite data. International Journal of Remote Sensing, 15: 583–596.
- Besag J (1986). On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society, Series B, 48: 259-302.

References

- Canada Centre for Remote Sensing (2003). Principles of remote sensing. Centre for Remote Imaging, Sensing and Processing (CRISP) National University of Singapore.
- Carfagna E, Gallego FJ (2005). Using remote sensing for agricultural statistics. International Statistical Review, 73: 389-404.
- Cressie N (1993). Statistics for spatial data. John Wiley & Sons, Inc., New York.
- Cumani A (1991) Edge detection in multispectral images. CVGIP: Graphical Models and Image Processing, 53: 40-51.
- Dorigo WA, Zurita-Milla R, de Wit AJW, Brazile J, Singh R, Schaepman ME (2007). A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. International Journal of Applied Earth Observation and Geoinformation, 9: 165-193.
- Everitt BS, Landau S, Leese M, Stahl D (2011). Cluster analysis. John Wiley & Sons, Ltd., Wiley, Chichester West Sussex.
- Fitzgerald GJ, Lesch SM, Barnes EM, Luckett WE (2006). Directed sampling using remote sensing with a response surface sampling design for site-specific agriculture. Computers and Electronics in Agriculture, 53: 98-112.
- Geman D, Geman S, Graffigne C, Dong P (1990). Boundary detection by constrained optimization. IEEE - Transaction on Pattern Analysis and Machine Intelligence, 12: 609-628.

References

- Geman S, Geman D (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE - Transaction on Pattern Analysis and Machine intelligence, 6: 721-741.
- Haining R, Arbia G (1993). Error propagation through map operations. Technometrics 35, 3: 293-305.
- Jensen JR (2004). Introductory digital image processing: A remote sensing perspective. Prentice Hall, New Jersey.
- Kirkpatrik S, Gelatt Jr CD, Vecchi MP (1983). Optimization by simulated annealing. Science, 220: 671-680.
- MacQueen JB (1967). Some methods for classification and analysis of multivariate observations. Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley University of California Press 1: 281-297.
- Marr D, Hildreth E (1980). Theory of edge detection. Proceedings of the Royal Society of London, Series B, Biological Sciences, 207: 187-217.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953). Equations of state calculations by fast computing machines. Journal of Chemical Physics, 21: 1087–1092.
- Muchoney DM, Haack BN (1994). Change detection for monitoring forest defoliation. Photogrammetric Engineering and Remote Sensing, 60: 1243 - 1251.

References

- Postiglione P, Andreano MS, Benedetti R (2013). Using constrained optimization for the identification of convergence clubs. Computational Economics, 42: 151-174.
- Renez AN, Ryerson RA (1999). Manual of remote sensing. 3rd ed., John Wiley & Sons, Inc., New York.
- Richards JA, Jia X (2006). Remote sensing digital image analysis. An Introduction. Springer Verlag, Berlin Heidelberg
- Ripley BD (1988). Statistical inference for spatial processes. Cambridge University Press, Cambridge.
- Strenski PN, Kirkpatrick S (1991). Analysis of finite length annealing schedules. Algorithmica, 6: 346–366.
- The World Bank (2011). Global strategy to improve agricultural and rural statistics. Report No. 56719-GLB, Washington, DC, USA.
- Thomson SJ, Zimba PV, Bryson CT, Alarcon-Calderon VJ (2005). Potential for remote sensing from agricultural aircraft using digital video. Applied Engineering in Agriculture, 21: 531−537.
- van Laarhoven PJM, Aarts EHL (1987). Simulated annealing: Theory and applications. Reidel Publishing Company, Dordrecht (Holland).

Setting up the frame

- The sampling frame has significant implications on the cost and the quality of any agricultural or non-agricultural survey.
- Defective sampling frames are a common source of nonsampling error, particularly under-coverage of important population sub-groups.
- (Särndal et al., 1992): The sampling frame is any material or device used to obtain observational access to the finite population of interest, and refers to the list of units in the survey population.
- It must be possible with the aid of the frame to identify and to select a sample in a way that respects a given probability sampling design and to establish contact with selected elements.
- A sampling frame could be a list of units (such as households or holdings), geographical areas, or any other materials (such as maps), and may also include information about each unit, such as their size, to help with the sample selection or survey estimation.

Setting up the frame

Main **properties** of a frame:

- the units of the frame are identified through a unique identifying code;
- all units, that are selected in a sample, should be contactable;
- the frame contains a set of additional information for each unit;
- each element in the population of interest is present in the frame (completeness), and only once (accuracy);
- there is no element in the frame that is out of the target population;
- frame data are up-to-date.

In **practice**: less than perfect situations, as several limitations can affect the quality of the frames. Ignoring or failing to properly identifying these limitations can lead to biased estimates, inflated variances, or even both.

Setting up the frame

FRAMES

LIST FRAMES

List of all the statistical units in the survey population.

Frame components: lists of addresses for holders or farmers.

Survey design:
list-sampling design

SPATIAL FRAMES

Complete and exhaustive list of non-overlapping geographic areas

Frame components: segments or portions of land

Survey design:
spatial sampling design

Setting up the frame

- In agricultural surveys, farms are traditionally sampled on a list (often a census) of farms.

- When a satisfactory list frame does not exist, and it would be too expensive or complex to create or when it is too expensive or complex to maintain it, a spatial frame may be the best alternative (Kennel 2008).

- Generally, agricultural censuses are made every 10 years so there may be a substantial difference between the sampling frame and the actual population at the date of the survey.

- **Good alternative**: Spatial frames.

- Sometimes spatial frames are used to improve an imperfect frame. For example, a survey of agricultural holdings might use a frame of telephone numbers supplemented by a spatial frame. The sample drawn from the telephone list will not cover holdings without telephone service. However, constructing the entire survey from a spatial frame may be too expensive. Thus some surveys use a spatial frame to improve a frame with known coverage defects.

Setting up the frame

ADVANTAGES OF SPATIAL FRAMES

- ***Complete coverage*** *of the target population*, non overlapping, no gaps between adjacent units, unbiasedness of the survey estimates with measurable and higher precision.

- ***Timeliness***: timely estimates on cultivated areas and on expected productions.

- ***Longevity***: once a spatial frame is constructed, it remains updated for a long time, only updates for land use changes.

- ***Versatility***: since sampled reporting units can be associated with a portion of land, a spatial frame can be used to collect data for multiple variables in one survey, e.g. crop surface, livestock, economic and environmental data.

Setting up the frame

- ***Objectivity*** of data collection and non-sampling errors reduced. The materials used for the survey and the information collected help to reduce non-sampling errors in interviews and are a good basis for data imputation for non-respondents.
- ***Reduced burden*** on farmers.
- They ***allow the use*** of spatially defined auxiliary information (e.g. GIS data and remote sensing).

Setting up the frame

DISADVANTAGES OF SPATIAL FRAMES

- ***Cost***: The setting up of the frame is usually a high cost activity (not so high for point frame), high sample selection costs above all in the case of units with identifiable physical boundaries, face-to-face interviews conducted by well-trained enumerators are very costly, and high-tech methods, qualified office staff, and statisticians are needed.
- ***Not suitable*** for cultivations with high spatial variability (scattered), and they have a limited precision of the estimates for small areas or highly concentrated land cover/ use classes.
- ***Cartographic material*** is required to construct the frame (such as maps, satellite images, aerial photos).
- ***Less efficient*** than a list frame as the only stratification is for land cover/use and for commodities on large farms or commodities that are rare (but spatially concentrated).
- ***Sensitive*** to the impact of outliers, and the estimates may be unstable.

Setting up the frame

LIST FRAMES

- Lists of holdings or holders' addresses.
- Derive from previous agricultural, housing or population censuses, from lists set up by political or administrative subdivisions, from farmer's associations or from other administrative data sources.
- They contain information on holding size, crops, livestock, and other characteristics.
- These types of information are often used to stratify the frame that greatly improves sampling efficiency.
- Many different ways of preparing a list of agricultural holdings.
- Local knowledge should be used to list not only farm households, but also holdings under different legal status, such as cooperatives, government farms, and enterprises.

Setting up the frame

- In the case of countries, which have and keep up-to-date land records (a cadastre), it may be easy to prepare a frame through reference to the land records. In the land records, the name of the holder is normally entered together with all the fields operated by him. Checking the internal consistency of the cadastre is essential. By covering each field of the study region, it should be possible to completely list the holders.

- In countries where rural reconstruction and development programs have been initiated, a list of households may be readily available, and this can be used to screen the farm households. In case no list of households or holdings is present, it will be necessary to prepare a new listing of households operating some land. The enumerator may be instructed to start from a fixed point of the village, and systematically number every house in which there may be more than one holding. Only some of the members of the household may be agricultural holders. The enumerator should proceed from house to house, listing households along with the information whether its members are agricultural holders.

Setting up the frame

- The list frames are not generally linked with the territory, except if the parcels of the households and farms are digitized. This is obviously a very expensive task.

- if this link exists, the operational sampling frame includes the geographic dimension of the units, such as farms and households that should be connected to the land cover and/or land use dimensions. This produces a series of benefits. Among the others, the link of the farm with its position in the territory ensures the assessment of the quality of self reported responses of farmers, and the use of these measurements for benchmarking. Furthermore, this association facilitates agro-environmental analysis. Remote sensing data can help in handling this problem by adding the geographical dimension to the list frames.

Setting up the frame

SPATIAL FRAMES

- In case of spatial units with segments that have recognizable physical boundaries, the frame is an ordered list of land areas, called frame units, with their assigned number of segments, such that they form a complete subdivision of the total land area of each land-use stratum, with no overlap. They provide a clear-cut means of identifying of each segment, and the number of segments assigned to each frame unit facilitates the probability sampling of segments.

- The preparation of such a frame is a very demanding issue.

- Up-to-date cartographic material (maps, satellite images, aerial photos) are required on which the land to be included can be visualized.

- The resolution or detail of the material must be sufficient to stratify according to intensity of land use and the subsequent subdivision of these land-use strata into frame units also with recognizable physical boundaries.

Setting up the frame

- Land-use strata and frame units are identified on satellite images or on a mosaic of aerial photographs and then transferred to topographic charts and measured.

- Frame units are constructed generally with maps on which the boundaries of the land-use strata have been transferred.

- In each land-use stratum, each frame unit must be measured and assigned to a target number of segments of approximately equal size. Then, the number of segments assigned to each frame unit is summed to provide the total number of segments in the stratum, and a sample of segments is selected from each land-use stratum. Each sample segment is constructed on small mosaics of aerial photography on which the boundary of the corresponding frame unit have been transferred. The selected sample segments are located on appropriate aerial photo enlargements used to control field data collection.

Setting up the frame

- In the case of spatial frames with regular polygons or with segments that coincide with the land of agricultural holdings, the frame construction is simpler than in the previous case.

- The construction of point frames is basically simpler as the generation of coordinates is made in an automatic way. The grid, produced by software, is then overlaid with a *point to polygon* map intersection in order to attribute to each point one and only one territorial nomenclature code, based on administrative divisions (e.g. NUTS) excluding points outside the study region (e.g. water, foreign States, etc.). Then, each point is overlaid on aerial photos or low-resolution satellite images so that a photo-interpreter can assign a stratum code.

- The frame is often overlaid with a digital elevation model to assign to each point, the elevation, which could be useful for cut-off sampling or stratification

Setting up the frame

- The reduction of **non-sampling errors** is a strategy often crucial for the success of a survey and of these errors probably the most influential are those deriving from the list used for the representation of the target population.

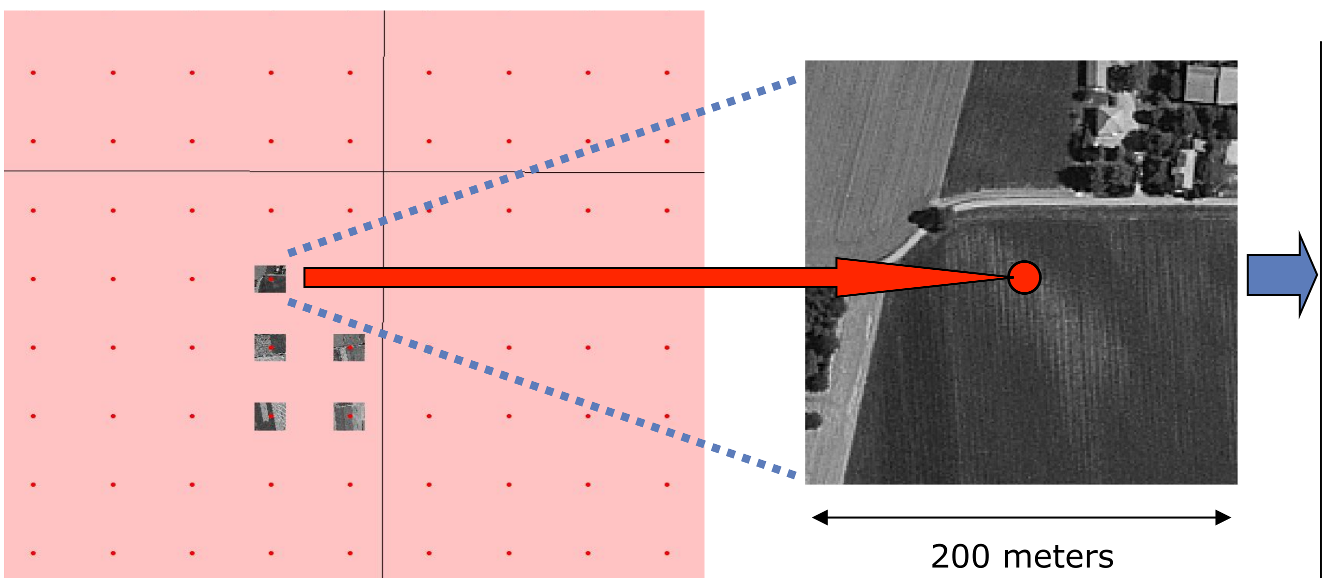
- Any difference between the frame and the target population implies a possible **bias** in the results whose weight will grow with the increase of the distance between the two populations.

Stratification

Square Grid, say 500m, super-imposed over geographical map

Photo interpretation

Stratification

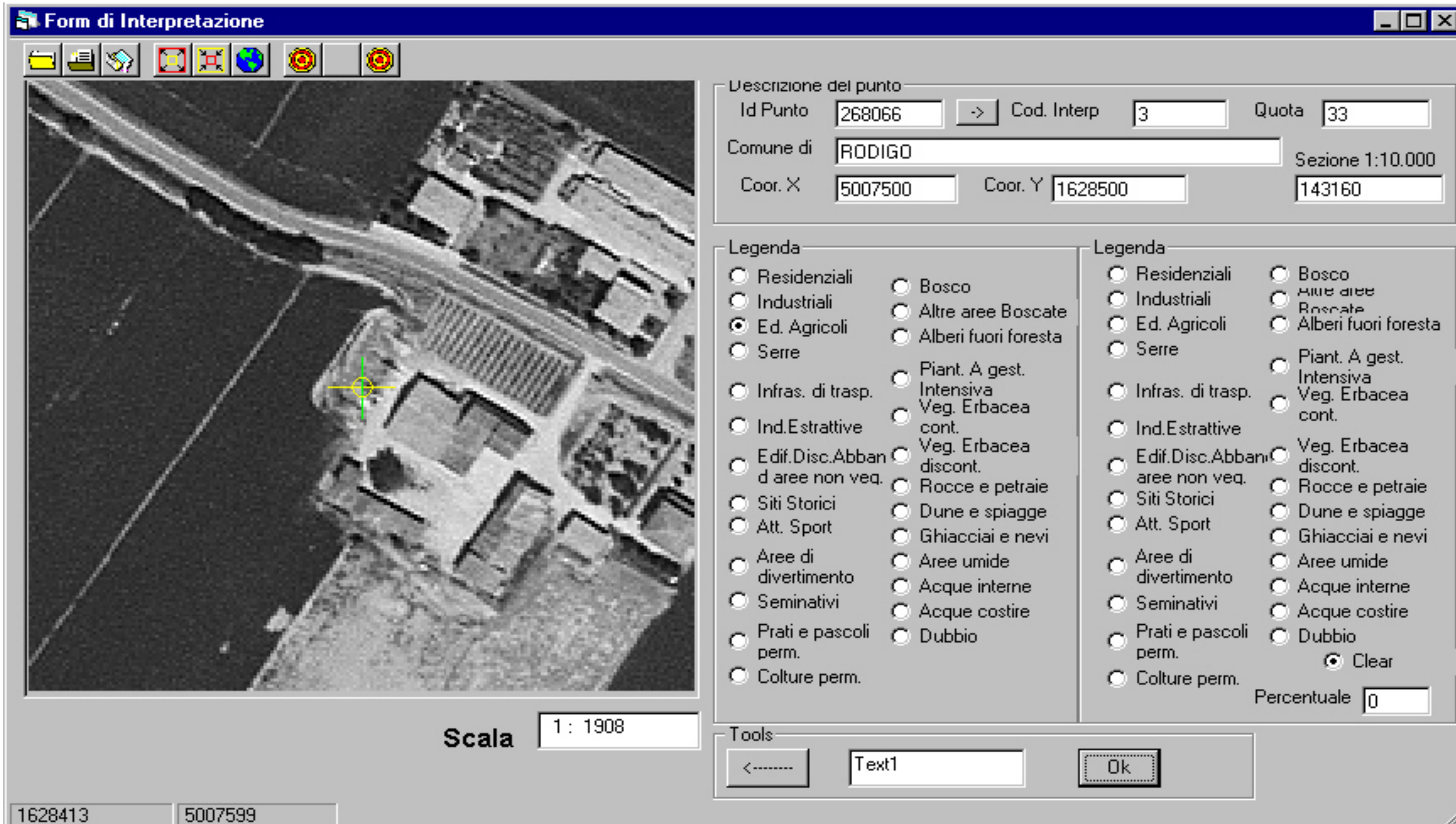


LAND USE
1................
2. **AGRICULTURAL AREA**
   - 2.1 ARABLE LAND
   - 2.2 PERMANENT MEADOWS AND PASTURES
   - 2.3 LAND UNDER PERMANENT CROPS

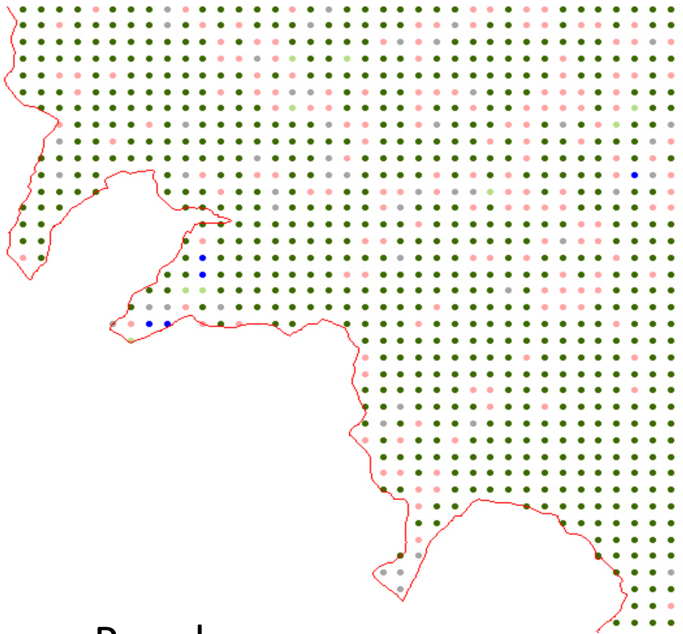
3..................
4................
5...................
6...........

A software to speed up this long and boring activity is very important

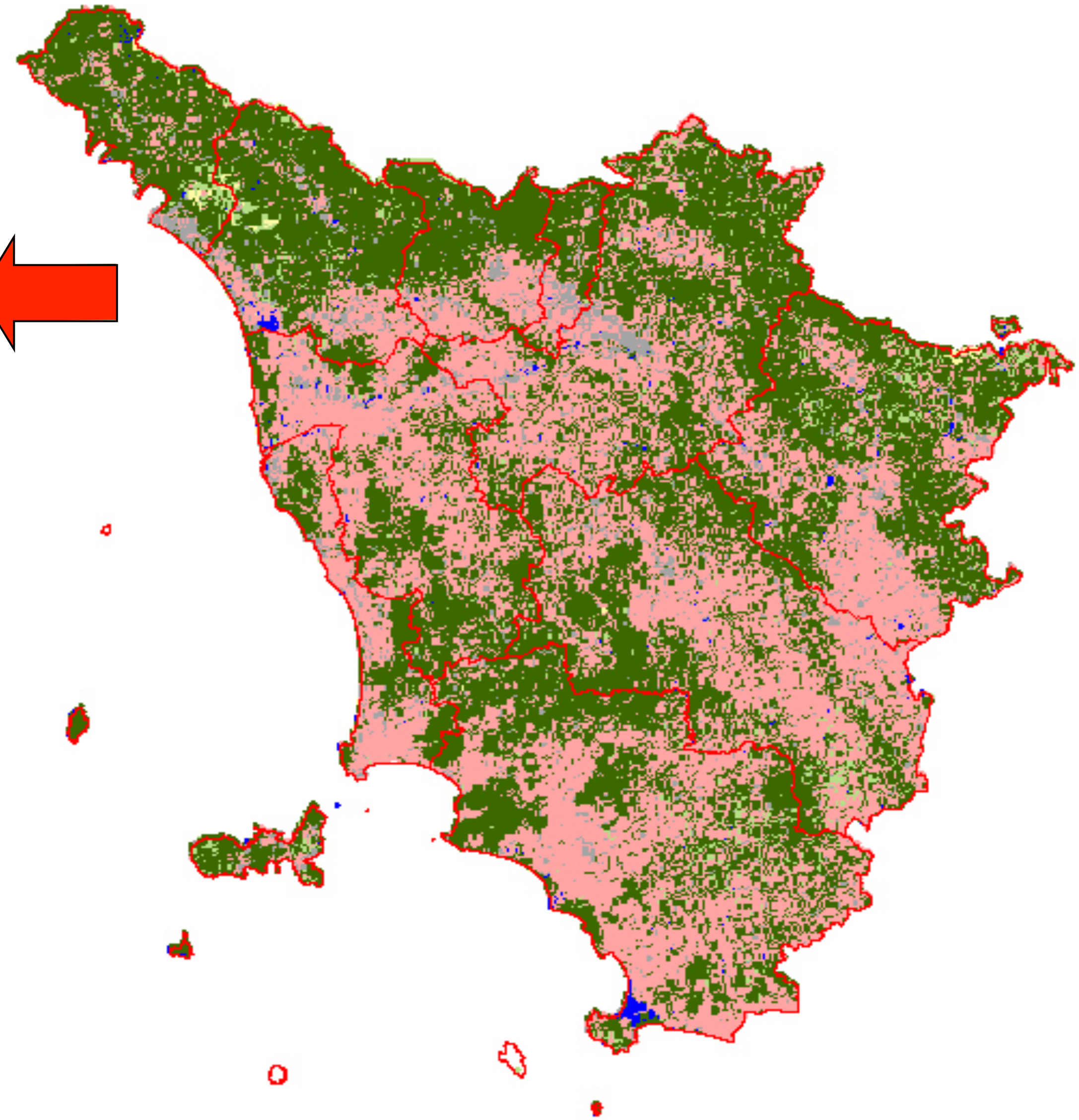
POPOLUS: photo- interpretetation software



Stratification



Popolus
- Artificial land
- Utilized Agricultural Area
- Woodland
- Grassland
- Bare land
- Wetlands/Water Areas



Region: Tuscany 91.901 points

Stratification

Interpreted Land Use

Domains

Regions

Districts

Smaller ?

Slope (Irrigated Crops)

Only a Threshold

Elevation

Usually Only a Threshold

Cut-Off Sampling

Setting up the frame

Spatial units: segments of land areas

Classical Land/Agric. Finite Populations

- REGULAR POLYGONS

  (set by the use of a squared grid of map coordinates: square segments whose end points are established by map coordinates)
- IRREGULAR POLYGONS

  (set by identifiable physical boundaries such as rivers, roads, etc.)
- POINTS

Other Populations in Environmental Studies

- LINES or TRANSECTS
- Continuous SURFACES
- ANIMALS (finite but unknown)

Setting up the frame

Advantages/disadvantages

1. REGULAR POLYGONS : The area frame construction/management and sample selection are similar but simpler than in 2; reasons of economy; better geometrical properties (overlay with satellite data); delays of the production of estimates

2. IRREGULAR POLYGONS: reasons of economy; limits should follow phisical boundaries thus it should be easier to identify the statistical unit than in 1 and 3; it usually requires selection ΠPS; better geometrical properties (overlay with satellite data); delays of the production of estimates

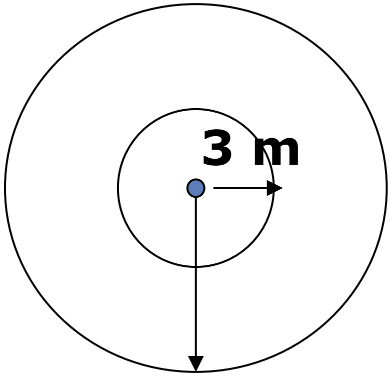
3. POINTS: The area frame construction (in particular its stratification) and sample selection is simpler than in 2 and the estimation methods are simpler than for 1 and 2 but usually less efficient given n; data collection and processing are easier than for 1 and 2; some survey variables are point-specific and cannot be measured over estended polygons

Setting up the frame

Point

Theoretical definition: no dimension but it may be defined as having a certain size for coherence with the observation rules or the location accuracy that can be achieved

Operational definition: circular portion of land



Regular polygon

Theoretical definition: regular dimension, all the same size

Operational definition: square portion of land

Irregular polygon

Theoretical definition: irregular dimension, varying

Operational definition: irregular portion of land

References

- Carfagna E (2001). Multiple frame sample surveys: advantages, disadvantages and requirements. Bulletin of the International Statistical Institute, 53rd Session, Proceedings, Book 1, Invited papers, pp. 271-274.
- Carfagna E (2013). Using satellite imagery and geo-referencing technology for building a master sampling frame. Proceedings of the 59th ISI World Statistics Congress, Hong Kong.
- Carfagna E, Carfagna A (2010). Alternative sampling frames and administrative data. What is the best data source for agricultural statistics? In: Agricultural survey methods, Benedetti R, Bee M, Espa G, Piersimoni F (eds), John Wiley & Sons Ltd, Chichester, UK, pp. 45-61.
- Carfagna E, Gallego FJ (2005). Using remote sensing for agricultural statistics. International Statistical Review, 73: 389-404.
- Colledge M (2004). OECD/UNESCAP/ADB. Workshop on assessing and improving statistical quality: measuring the non-observed economy. Bangkok, 11-14 May 2004.
- Cotter J, Davies C, Nealon J, Roberts R (2010). Area frame design for agricultural surveys. In: Agricultural survey methods, Benedetti R, Bee M, Espa G, Piersimoni F (eds), John Wiley & Sons Ltd, Chichester, UK, pp. 169-192.
- FAO (1989). Sampling Methods for Agricultural Surveys. Rome.
- FAO (1995). Fao Statistical Development Series No. 6 - Conducting Agricultural Censuses And Surveys.

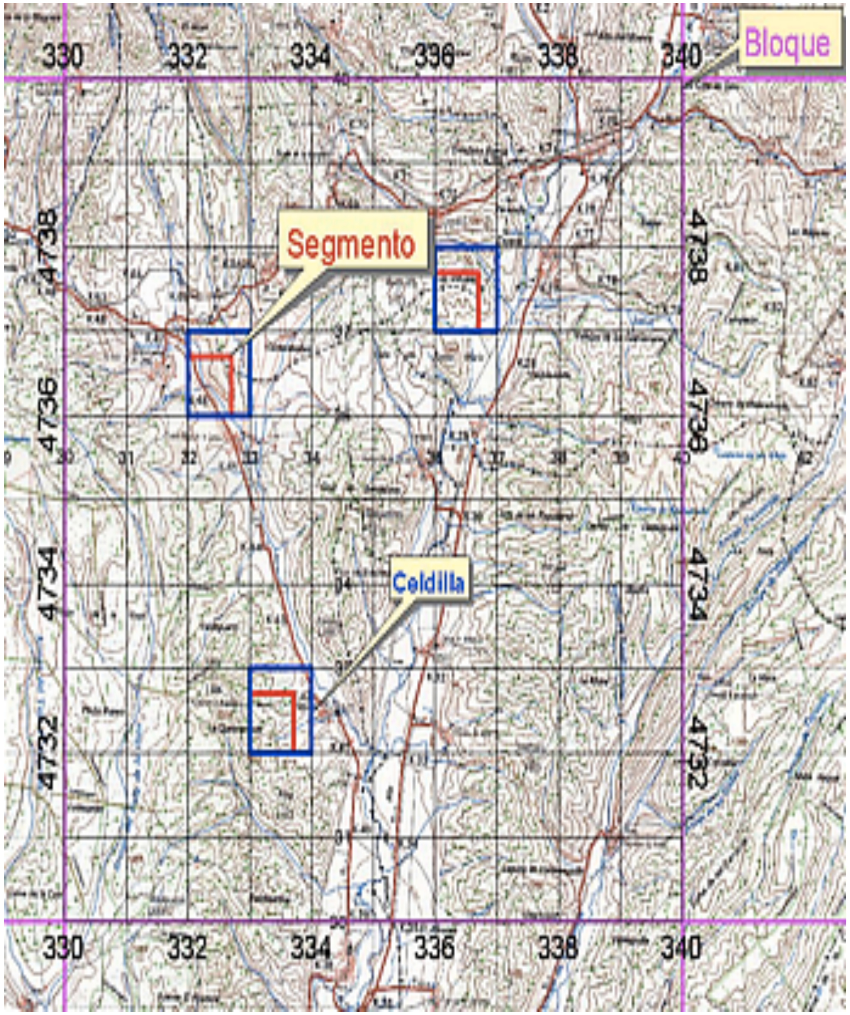
References

- FAO (1996). Multiple frame agricultural surveys, Volume 1. Current surveys based on area and list sampling methods. FAO Statistical Development Series 7. Rome, FAO.
- FAO (1998). Multiple frame agricultural surveys, Volume 2. Agricultural survey programmes based on area frame or dual frame sample designs. FAO statistical development series 10.
- FAO (2005). A System of Integrated Agricultural Censuses and Surveys. World programme for the census of agriculture 2010. FAO statistical development series 11.
- Fecso R, Tortora RD, Vogel FA (1986). Sampling frames for agriculture in the United States. Journal of Official Statistics, 2: 279–292.
- Ferraz C (2013). Improving methods for linking area frames with list frames: literature review preliminary results, Global Strategy, First Scientific Advisory Committee meeting, FAO, July 2013.
- Flores LA (2013). Identifying the most appropriate master frame for an integrated survey, Global Strategy, First Scientific Advisory Committee meeting, FAO, July 2013.
- Gallego FJ (1999). Crop area estimation in the MARS project. Conference on ten years of the MARS project, Brussels, April 1999.
- Kennel T (2008). Area frame. In: Lavrakas P J (ed), Encyclopedia of Survey Research Methods, Sage , pp. 31-33.
- Nusser SM, House CC (2009). Sampling, data collection, and estimation in agricultural surveys. In: Sample surveys: design, methods and applications, Pfeffermann D, Rao CR (eds), vol. 29A, Elsevier, The Netherlands, pp. 471-486.

ESYRCE (Spain)

- Square segments on a regular grid of 10km x 10km (block), cells of 1km x 1km: 3 cells for every block with the same relative position (systematic sampling).
- Working unit: segment (700m x 700m) in the SW corner of the cell
- National stratification for land cover: 6+ segments in highly cultivated blocks
- Visit of the segments: the surveyor, with the help of maps, designs hand made parcels boundaries overlaied on aerial photos and fills a questionnaire

ESYRCE (Spain)



LUCAS (Europe)

LUCAS started as a pilot survey across a limited number of EU Member States to provide early crop estimates for the European Commission. The first survey was held in 2001 and in 2009 the focus of the survey changed from a merely agricultural to a broadly agro-environmental one and it *focuses on state, dynamics and changes*

Aims

- Produce harmonised and comparable land cover/use statistics at the European level through a standard survey methodology (for environmental and economic projects linked to the sustainable use of resources and to climate change)
- Monitor changes in land cover and land use.
- Analyse the interactions between agriculture, environment and countryside (such as irrigation, land management ect.).
- Analyse soil quality and to provide ground truth and calibration for many space-borne information collection activities.

LUCAS (Europe)

Land cover refers to the bio-physical coverage of land (for example, natural areas, forests, buildings and roads or lakes).
Land use refers to the socioeconomic use that is made of land (for example, agriculture, commerce, residential use or recreation); at any one place, there may be multiple and alternative land uses.

LUCAS (Europe)

Artificial land

Cropland

Woodland

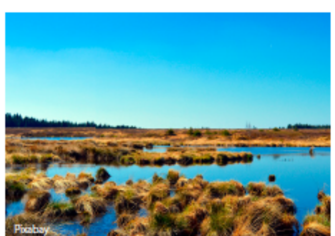
Shrubland

Grassland

Bare land

Water areas

Wetlands



Primary sector:

(for example, agriculture and forestry)

Secondary sector

(industry)

Tertiary sector

(services)

Other uses

(for example, residential use and abandoned areas)

LUCAS (Europe)

Some features

- Flexible modules dealing with more specific themes, such as soil and biodiversity.
- Minumum statistical burden (the surveyors do not ask any questions to the local residents, land owners and users).
- In 2018 survey 28 EU countries involved and about 240.000 geo-referenced points visited by 750 field surveyors on the spot and 98.000 photo interpreted.
- Possibility to track the changes in land cover and land use over time (many points are in common in the ≠ surveys).
- Results are reliable at EU level down to regions and provinces depending on the country.

LUCAS (Europe)

- **Sampling strategy**: *2-phase stratified area frame sampling of points covering the entire EU territory*.
- **Aims**:
  - obtaining a reasonable balance among the strata (the agricultural strata were over-represented in the previous rounds of the survey) since the focus of the survey changed from a merely agricultural to a broadly agro- environmental one;
  - providing reliable estimates at a geographical level more detailed than the EU one for the most relevant land cover classes in the various countries;
  - gathering longitudinal data on land cover and land use to monitor their changes (for the countries involved in the 2006 round).
- **Harmonised nomenclature**: applicable in all EU Member States
- **Harmonised data collection method**: in all countries (Field Survey according to well defined instructions)

LUCAS (Europe)

- LUCAS is a two phase sample survey.
- First phase sample: systematic sample with points spaced 2 km., i.e. standard 2 km grid with in total around 1 million points all over the EU.
- Each point of the first phase sample was photo-interpreted and assigned to one of the 7 pre-defined land cover strata.
- The selection of points (second phase) is done on the basis of stratification information and they are classified by field visits according to the full land cover and land use nomenclature.
- The stratified sample was selected independently in each NUTS2 region fixing precision targets on the estimates of the main land cover classes.
- The statistical unit is a portion of land of circular shape and a conventional dimension of 1.5 meter radius (extended to 20 meters in specific cases).

LUCAS (Europe)

- Nomenclature LUCAS: Land Cover

A10	Built-up areas
A20	Artificial non built-up areas
B10	Cereals (+ triticale)
B20	Root crops
B30	Non permanent industrial crops
B40	Dry pulses, vegetables and flowers
B50	Fodder crops
B70	Fruit trees & berries
B80	Other Permanent Crops
C10	Broadleaved and evergreen woodland
C20	Coniferous woodland
C30	Mixed woodland
D10	Shrubland with sparse tree cover
D20	Shrubland without tree cover

E10	Grassland with sparse tree/shrub cover
E20	Grassland without tree cover
E30	Spontaneous vegetation
F00	Bare Land
G10	Inland water bodies
G20	Inland running water
G30	Coastal water bodies
G50	Glacier, permanent snow
H10	Inland wetlands
H20	Coastal wetlands

8 categories

24 subclasses

76 classes

LUCAS (Europe)

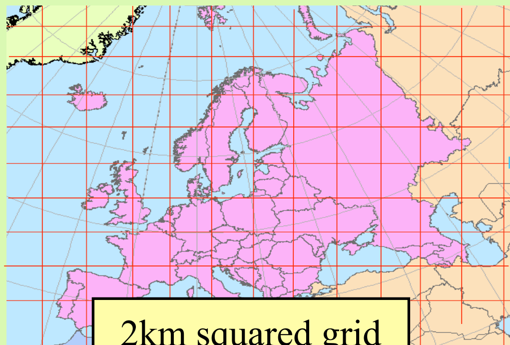
Nomenclature LUCAS: Land Use

Composed by 15 categories and 34 classes.

U110	Agriculture (+ Kitchen garden + Fallow land)
U120	Forestry
U130	Fishing
U140	Mining, Quarrying
U150	Hunting
U210	Energy production
U220	Industry & Manufacturing
U310	Transport, communication, …
U320	Water & waste treatment
U330	Construction
U340	Commerce, Finance, Business
U350	Community Services
U360	Recreation, Leisure, Sport
U370	Residential
U400	Unused

LUCAS (Europe)

First phase sample for stratification: orthophoto interpretation



2km squared grid

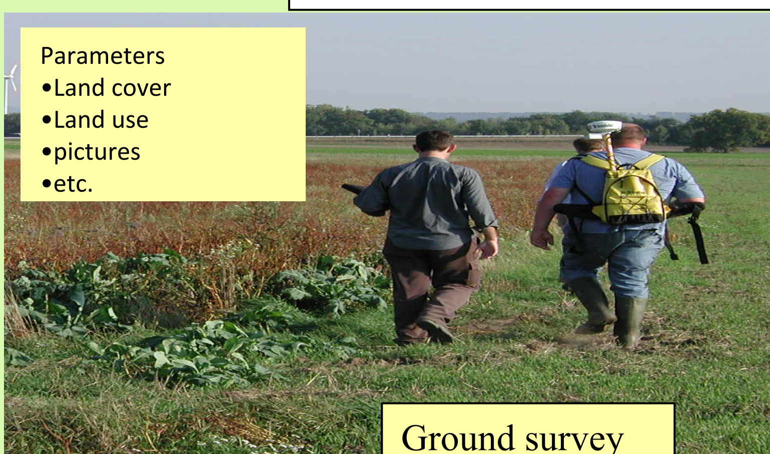
1 100 000 points

LAND COVER classes

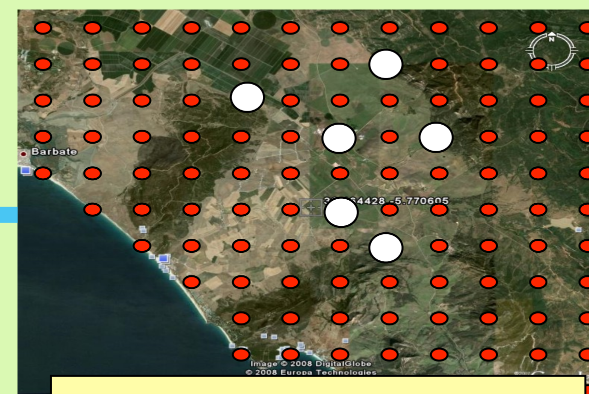
1 ARABLE LAND
2 PERMANENT CROPS
3 GRASSLAND
4 WOODED AREAS AND SHRUBLAND
5 BARE LAND, RARE VEGET.
6 ARTIFICIAL LAND
7 WATER

Second phase sample: in-situ data collection

Parameters
- Land cover
- Land use
- pictures
- etc.

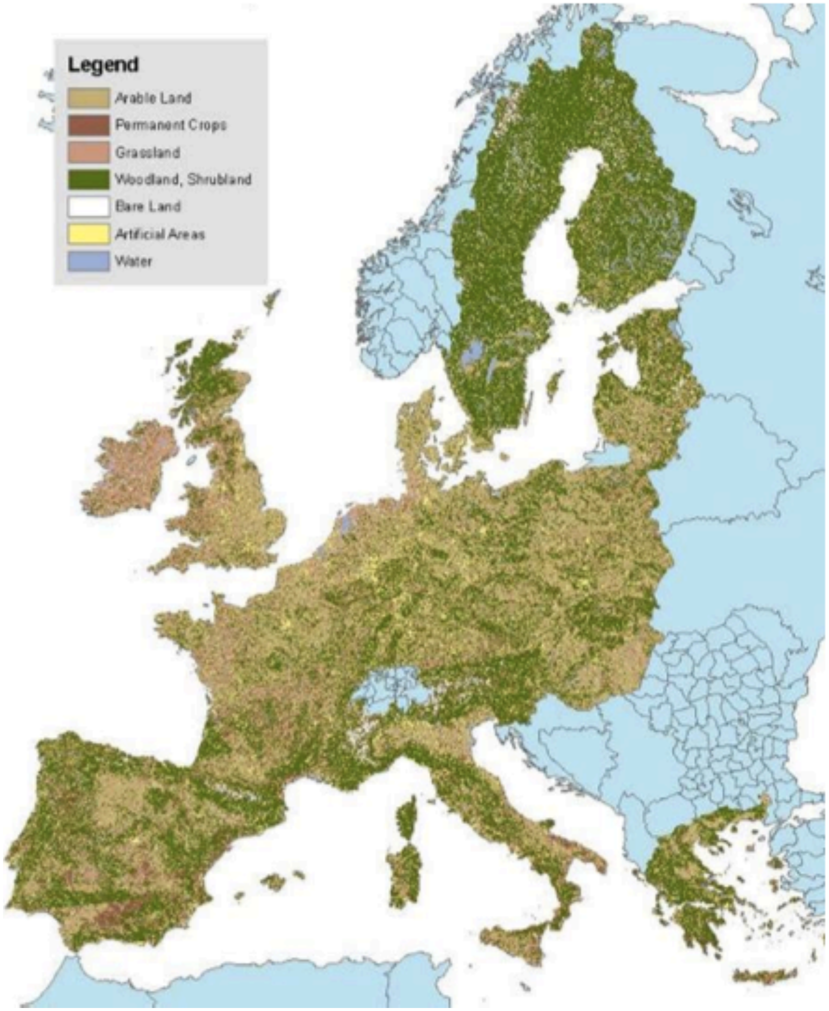


Ground survey



Sample of around 265,000 pts

LUCAS (Europe)



Stratification result:

Photo- interpretation of more than 1 million points

- arable land
- permanent crops
- grassland
- wood/shrub land
- bare land, low/rare vegetation
- artificial land
- water

LUCAS (Europe)

The observation **Point**

The extended observation **window**

LUCAS (Europe)

- Sampling drawing:

- Selection of points: criteria
  - LUCAS sample points are included as much as possible (panel approach)
  - Points above 1000m and very difficult to access are excluded
  - Strategy to reduce autocorrelation among points

LUCAS (Europe)

- The heterogeneity of land cover and the presence of linear features such as walls, hedges, roads, railways or irrigation channels are two key elements characterising landscape structures.
- Purpose: identify the typologies of countries with a different degree of land cover diversity and landscape patterns (combination of natural and human features) with diversity indicators
- Straight line of 250m in Eastern direction from the observation point

Land cover features	*Durum wheat*	*Grassland with trees*	*Wood margin*	*Maize*	*Oak tree line*	*Maize*	*Road (>3 m)*	*Ditch*	*Mixed forest*
Codification in field form	*B12*	*E10*	*15*	*B16*	*11*	*B16*	*62*	*31*	*C30*

Observation of:

- Land cover transitions
- Linear features

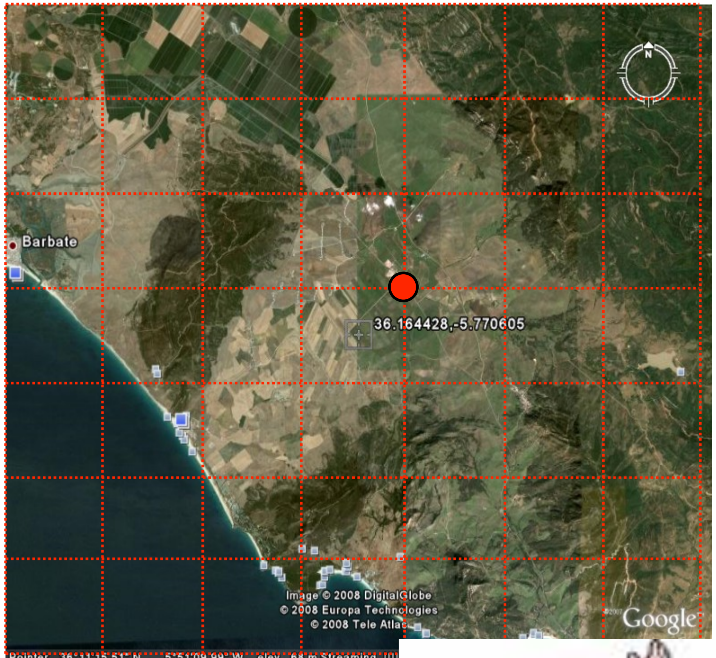
are recorded in the sequence of their appearance

LUCAS (Europe)

"LUCAS Soil 2018": on 10 % of the points, 20 000 random top soil samples were collected by the LUCAS surveyors at points of their visit and later analysed in a soil laboratory: to measure the quantity of organic carbon in soil, to update the European soil maps, to assess environmental factors and to validate soil models

- SOIL LABORATORY ANALYSIS
  - Coarse fragments
  - Particle size distribution (FAO, 1990a)
    - Clay content
    - Silt Content
    - Sand Content
  - pH(CaCl2)
  - pH(H2O)
  - Organic carbon

LUCAS (Europe)



Microdata - Parameters:

- Land cover (1 & 2)
- Land use (1 & 2)
- Land Cover percentage
- Area size
- Height of trees
- Width of features
- Transect (250 m to east) with Land Cover and linear features
- Land management (grazing)
- Water management

SOIL SAMPLE

TERUTI (France)

- The survey began in 1946 with the control of agricultural land based on cadastral plans.
- In 1962, aerial photography is introduced, not as support of the survey but as a tool updating cadastral plans.
- In 2005 technical progress on cartographic documents and georeferencing helped in redefining the sampling of Teruti points and the beginning of Lucas encouraged a redesign of Teruti to ensure a coherence on the nomenclature, on the method of observation and on the sample design.
- Since 2012, the Teruti points located on agricultural land are filled in from applicants' surface administrative declarations.

TERUTI (France)

Purposes

• to know the different classes of land cover/use for the whole territory (agricultural, environmental and urbanized), yearly, at different geographical levels;
• to follow and quantify the land cover/use and landscape structure changes;
• data source for other studies (environmental indicators, etc);
• to increase the sample (number of segments or number of points per segment).

TERUTI (France)

- TER-UTI uses a non-stratified two-stage sampling scheme, with points grouped in PSU or cluster or segments to reduce trips:
- first stage unit is the segment: square area of 1,5 kmX1,5 km;
- second stage unit is the point: land area, 3 m radius circle (general) or 40 m (heterogeneus land cover). Distance inside segment: 300 m.
- Two stage sampling scheme: almost the same precision as one stage but less expensive.
- Sampling basis: grid 3 kmX3 km. Every intersection is a segment.
- Sistematic sampling of the segments: 4 subsamples (replicates)
- “4”: grid 3 kmX3 km
- “1”: part of “4”, LUCAS 2003 points
- “2”: part of “4”, complement of “1”, grid 6 kmX6 km
- “3”: doubling of the segments number of "2" + "1"

TERUTI (France)

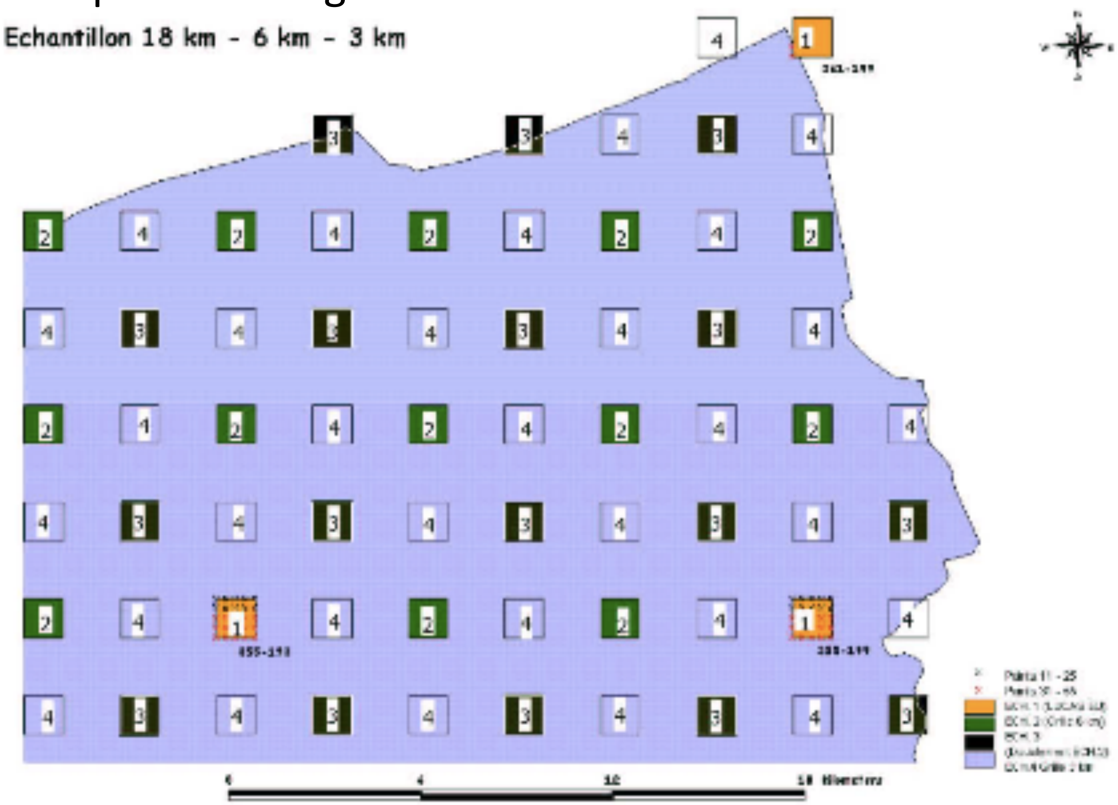
Sous échantillon	1 (18 km)	1+2 (6 km)	1+2+3 (6 km double)	1+2+3+4 (maître 3 km)
France	**1 764**	**15 896**	**31 655**	**64 492**

Replicate 2: segments 1+2

Replicate 3: segments 1+2+3

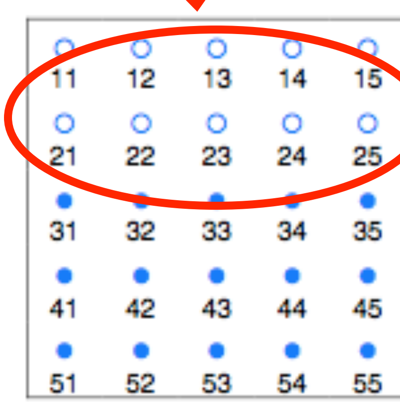
Replicate 4: segments 1+2+3+4

Goal yearly survey: 33.013 segments of 10 points
Good results for sub-regional level for most important and/ or homogeneous land covers



LUCAS

Pos. ext.



JAS (Usa)

Area sampling frames: vehicle for conducting surveys to gather information regarding crop acreage, cost of production, farm expenditures, grain yield and production, livestock inventories and other agricultural items.

Key steps:

- Basic stratification: (1) divide the land into land-use strata such as intensively cultivated land, urban areas and range land (2) divide each land-use stratum into substrata by grouping areas that are agriculturally similar.
- Multi-step sampling: (1) divide the land into larger sampling units called first-step or PSU; (2) select a sample of PSUs and then delineating the segments only for these PSUs; and (3) select a sample of segments from the selected PSUs.

JAS (Usa)

- The major area frame survey conducted by the NASS is the June Agricultural Survey (JAS): area frame estimates for crop acreages and livestock inventories.
- Identification of each segment on aerial photos, visit of the segment, interview of each person who operates land inside the boundaries of the selected segments.
- With the respondent's assistance, field boundaries are identified on the photography and the acreage and crop type reported for each field in the segment.
- Counts of livestock within each sample segment are also obtained.
- This area frame information is subsequently used to provide state, regional and national estimates for crop acreages, livestock inventories and other agricultural items.
- 49 states with approximately 11 000 segments being visited every year.

JAS (Usa)

Goals:

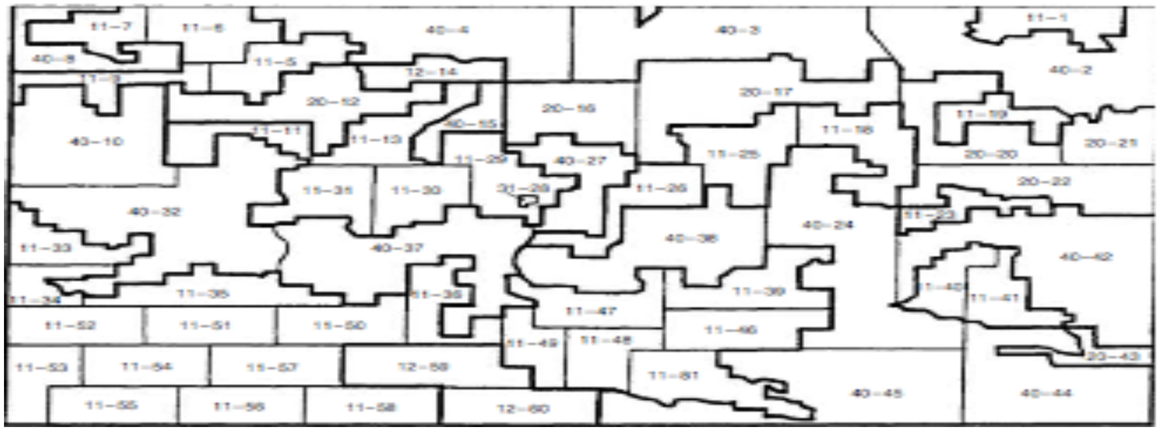
- *Acreage estimates for major commodities, livestock inventories and economic data*
- *Measure the incompleteness of the list*
- *Ground truth for remotely sensed crop acreage estimates (*what crop spectral signatures from the satellite represent)
- *Follow-on surveys (*Objective Yield Survey, Agricultural Coverage Evaluation Survey (ACES))

JAS (Usa)

- PSU and segment sizes: segment size varies from stratum to stratum and state to state and depend on Availability of boundaries (highways, roads, railroads, rivers,…), the minimization of sampling variability and data collection costs (typical segment about 1-2 square miles). PSU sizes are based on the segment size and contain six to eight segments.
- Every state has a sampling frame, where several land-use strata are common to all frames, including cultivated land, ag-urban, urban, and non-agricultural land. The cultivated land is divided into several strata based on the distribution of cultivation in the state.

JAS (Usa)

PSU ordering in a serpentine manner



Further level of stratification: each land-use stratum is further divided into substrata by grouping areas that are agriculturally similar i.e divide the population of sampling units within each stratum equally into categories (substrata). Sampling units are placed into substrata based on likeness of agricultural content and, to a certain extent, location. Sub-stratification activities include ordering the PSUs, ordering the counties, calculating the number of sampling units in the strata, determining the number of substrata, and placing the sampling units into substrata.

The JAS uses a sample comprised of designated land areas (segments) selected from this stratification.

JAS (Usa)

Within each substratum, the land is divided into primary sampling units (PSUs). A sample of PSUs is selected and smaller, similar-sized segments of land are delineated within these selected PSUs. Finally, one segment is randomly selected from each selected PSU to be fully enumerated

Each segment is outlined on an aerial photo (red) that is provided to the appropriate field enumerator. Through field enumeration, a segment is divided into tracts of land, each representing a unique land operating arrangement (blue). An area screening form is completed for all tracts within a segment and contains screening questions that determine whether or not each tract has agricultural activity.

JAS (Usa)

- Replicated sampling: selection of a number of independent subsamples or replicates from the same population using the same selection procedure for each replicate. Each replicate is therefore an unbiased representation of the population.
- Replicate: simple random sample of one land area (segment) selected from each substratum within a land-use stratum.
- Reasons: 5 years sample rotation, the total number of segments rotated each year is approximately 3.000 (reduce burden), Methodology research, Quality assurance, variance estimation, easy sample management and evaluate rotation effects.

JAS (Usa)

- Multivariate optimum allocation for the estimates at the state level.
- Unequal selection probabilities (PPS) is restricted to the non-agricultural stratum and open land strata (adequate boundaries are not available). In all other land-use strata equal probability of selection is used. About 96% of the approximately 11.000 segments in the area frame sample are selected based on the equal probability of selection method.

JAS (Usa)

Summary

- The June Area Survey (JAS) collects information on U.S. crops, livestock, grain storage capacity and type and size of farms.
- land is divided into homogeneous strata, such as intensively cultivated land, urban areas and range land
- Each land-use stratum is further divided into substrata by grouping areas that are agriculturally similar.
- Within each substratum, the land is divided into primary sampling units (PSUs).
- A sample of PSUs is selected and smaller, similar-sized segments of land are delineated within these selected PSUs.
- one segment is randomly selected from each selected PSU to be fully enumerated. Through in-person canvassing, field interviewers divide all of the land in the selected segments into tracts, where each tract represents a unique land operating arrangement. Each tract is screened.

AGRIT (Italy)

AGRIT is a survey projected to obtain areas estimates of the main crops according to the LUCAS nomenclature of land cover valid at the European level. The survey covers the entire area of the Italian territory.

The adopted sampling design is a stratified two-phase sampling:

1. the reference frame is defined: a regular grid is overlayed to the Italian territory resulting in N points that represent an aligned spatial systematic sample. Using aerial photos each point is then classified according to a land use hierarchical nomenclature with a total of 25 entries;

2. the final sample is selected: in each Italian province is selected a sample stratified by class of land use for a total of about 150.000 sampled points.

AGRIT (Italy)

- Integration of data collected in ground samples and remote sensing data
- List: set of points (point frame) which exhaustively covers the territory. Each point is provided with an operational dimension area: a 3 m radius circle centered on the reference point (thus with a total of about 30 m^2); in the presence of association between different land uses, the observation area is extended to about 700 m^2 by increasing the radius to 15 meters.
- Design: stratified in 2 (*3*) phases
- First-phase sample: aligned spatial systematic selection of geographical units on a regular grid of points (N about 1.2 million), with a spacing of 500 m
- Second-phase sample: stratification of the first-phase sampled units by province and 6 land cover classes (aerial photo interpretation); All points that show agricultural activity compose the reference population for the selection of the second-phase sample

AGRIT (Italy)

- From these strata and for each of the 103 Italian provinces a random second-phase subsample is extracted. Overall, approximately 150.000 points are collected on the ground.
- Aim: estimates with a predetermined sample error for a set of variables (i.e. areas of different crops)
- Bethel algorithm for stratified sampling and Horvitz–Thompson estimator
- *Third-phase* sample: yield estimates; Bethel's algorithm by using the variances of the yield of the previous AGRIT surveys. The sample size: about 60.000 points.

References

- Arroway P, Abreu DA, Lamas A C, Lopiano KK, Young LY (2010). An alternate approach to assessing misclassification in JAS. Proceedings of the Section on Survey Research Methods JSM 2010, American Statistical Association, Alexandria, VA.
- Benedetti R, Bee M, Espa G, Piersimoni F (2010). Agricultural survey methods. John Wiley & Sons Ltd, Chichester, UK.
- Benedetti R, Piersimoni F, Postiglione P (2012). Statistics in agricultural production. In: Methods and applications of statistics in the social and behavioral sciences, Balakrishnan N (ed), John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 357-373.
- Carfagna E, Gallego FJ (2005). Using remote sensing for agricultural statistics. International Statistical Review, 73: 389-404.
- Cotter J, Davies C, Nealon J, Roberts R (2010). Area frame design for agricultural surveys. In: Agricultural survey methods, Benedetti R, Bee M, Espa G, Piersimoni F (eds), John Wiley & Sons Ltd, Chichester, UK, pp. 169-192.
- EEA (2007). CLC2006 technical guidelines, EEA Technical report, No 17. (http://www.eea.europa.eu/publications/technical_report_2007_17)
- Everaers P (2010). The present state of agricultural statistics in developed countries: situation and challenges. In: Agricultural survey methods, Benedetti R, Bee M, Espa G, Piersimoni F (eds), John Wiley & Sons Ltd, Chichester, UK, pp. 1–24.
- EUROSTAT (2000). Manual of concepts on land cover and land use information systems. Theme 5: Agriculture and Fisheries: Methods and Nomenclatures. Office for Official Publications of the European Communities, Luxembourg.

References

- Eurostat (2019). https://ec.europa.eu/eurostat/statistics-explained/index.php/LUCAS_-_Land_use_and_land_cover_survey
- FAO (1996). Multiple frame agricultural surveys. Vol. 1: Current surveys based on area and list sampling methods. Statistical Development Series n.7.
- FAO (1998). Multiple frame agricultural surveys. Vol. 2: Agricultural survey programs based on area frame or dual frame (area and list) sample designs. Statistical Development Series n. 7.
- Gallego J, Delincè J (2010). The European land use and cover area-frame statistical survey, In: Agricultural survey methods, Benedetti R, Bee M, Espa G, Piersimoni F (eds), John Wiley & Sons Ltd, Chichester, UK, pp. 151-168.
- Haining RP (2003). Spatial data analysis: theory and practice. Cambridge University Press.
- Jacques P, Gallego FJ (2005). The LUCAS project The new methodology in the 2005/2006 surveys. Agri-environment workshop, Belgirate, September 2005.
- Lamas AC, Abreu DA, Arroway P, Lopiano KK, Young LJ (2010). Modeling misclassification in the June Area Survey. Proceedings of the Section on Survey Research Methods JSM 2010, American Statistical Association, Alexandria, VA.

References

- Nusser SM, House CC (2009). Sampling, data collection, and estimation in agricultural surveys. In: Sample surveys: design, methods and applications, Pfeffermann D, Rao CR (eds), vol. 29A, Elsevier, The Netherlands, pp. 471-486.
- Openshaw S (1977). A geographical solution to scale and aggregation problems in region–building, partitioning, and spatial modelling. Transactions of the Institut of British Geographers, New Series 2: 459–472.
- Openshaw S (1984). The Modifiable Areal Unit Problem, CATMOG 38, Geo Abstracts, Norwich.
- Postiglione P, Benedetti R, Piersimoni F (2010). Spatial prediction of agricultural crop yield. In: Agricultural survey methods, Benedetti R, Bee M, Espa G, Piersimoni F (eds), John Wiley & Sons Ltd, Chichester, UK, pp. 369– 387.
- Vidal C, Marquer P (2002). Changes in land cover and in land use. 2. Some findings. Statistics in Focus, Agriculture and Fisheries, Theme n.5, Eurostat.

Thank you for the attention!